

5610 12

Intelligent Pilot Aids for Flight Re-Planning in Emergencies

**Annual Performance Report
Grant NAG 1-01055**

**submitted to
NASA Langley Research Center
Hampton VA 23681-2199**

**for the period
April 1, 2001 – March 31, 2002.**

Attention: Anna Trujillo

**Amy R. Pritchett, Sci.D., Principal Investigator
Schools of Industrial and Systems Engineering and Aerospace Engineering
Georgia Institute of Technology
Atlanta, GA 30332-0205
(Tel) 404-894-0199
(Fax) 404-894-2301
Amy.Pritchett@isye.gatech.edu**

February 2002

This report covers April 1, 2001 to March 31st, 2002. During this time, experimental studies were conducted with pilots to investigate the attributes of automation that would be appropriate for aiding pilots in emergencies. The specific focus of this year was on methods of mitigating automation brittleness. Brittleness occurs when the automatic system is used in circumstances it was not designed for, causing it to choose an incorrect action or make an inaccurate decision for the situation (Billings, 1997). Brittleness is impossible to avoid since it is impossible to predict every potential situation the automatic system will be exposed to over its life. However, operators are always ultimately responsible for the actions and decisions of the automation they are monitoring or using, which means they must evaluate the automation's decisions and actions for accuracy. As has been pointed out, this is a difficult thing for human operators to do. There have been various suggestions as to how to aid operators with this evaluation. In the study described in this report we studied how presentation of contextual information about an automatic system's decision might impact the ability of the human operators to evaluate that decision.

This study focused on the planning of emergency descents. Fortunately, emergencies (e.g., mechanical or electrical malfunction, on-board fire, and medical emergency) happen quite rarely. However, they can be catastrophic when they do. For all predictable or conceivable emergencies, pilots have emergency procedures that they are trained on, but those procedures often end with "determine suitable airport and land as quickly as possible." Planning an emergency descent to an unplanned airport is a difficult task, particularly under the time pressures of an emergency (Pritchett, Nix, & Ockerman, 2001). Automatic decision aids could be very efficient at the task of determining an appropriate airport and calculating an optimal trajectory to that airport. This information could be conveyed to the pilot through an emergency descent procedure listing all of the actions necessary to safely land the plane. However, there is still the potential problem of brittleness. This study examined the impact of contextual information (Ockerman & Pritchett, 2000) in presentations of emergency descent procedures to see if they might impact the pilot's evaluation of the feasibility of the presented procedure. The study and its results are described in detail below.

Method

Participants and Apparatus

The participants of this investigation are current airline pilots. A total of 32 pilots participated in this study, with 28 of them choosing to provide demographic data. Those pilots have an average of 11,500 flight hours with just over 4000 hours in glass cockpits. Twenty-two of the participants are captains, five are first officers, and one was neither. The eight emergency descent scenarios they evaluated were presented on paper and consisted of 6 items: (1) a description of the emergency that has occurred along with a display of the current primary flight display and navigation display, (2) an enroute map for the new airport, (3) an approach plate for the new airport, (4) a STAR chart for the new airport, (5) horizontal and vertical map displays of the suggested descent path and (6) a text procedure for the suggested descent.

Procedure

The pilots were told that they were the captains of a glass cockpit commercial jet and that an emergency had occurred which required that they land at a different airport than originally planned. The scenarios' emergencies may or may not have affected the performance of the plane but did not have terrain or traffic conflicts. Each scenario consisted of a written description of the emergency and the current location of the plane. To replicate the time-criticality of emergency situations, the pilots were given 3 minutes to evaluate each emergency descent

procedure and record their response on the questionnaire. The pilots categorized each flight procedure as one they would be comfortable flying or one they would not be comfortable flying, and explained why or why not. They also provided their confidence in their response as a percentage.

Design

This experiment used a 2⁴ factorial design. The first factor was the condition of the aircraft (performance altered [PA] or not [NPA]), the second factor was the accuracy of the procedure, the third factor was the structure used in the presentation of the procedure, and the fourth factor was the presence of rationale (i.e., explanations).

Performance of the aircraft in each scenario was either altered in some way [PA] (e.g., lost engine or loose aileron) or was not [NPA] (e.g., sick passenger). Determining the future flight trajectory of a performance-altered aircraft is more difficult due to its unpredictable nature and inexperience with an aircraft in that particular condition.

Half of the eight emergency descent procedures were inaccurate. The inaccuracies were of two types. In one type of inaccuracy the graphic map display accompanying the procedure was redrawn to show a much tighter turn radius than feasible for the aircraft's speed and configuration at that point in the descent. In the other type the graphic vertical profile was altered to show an infeasible glide slope intercept, i.e., where the aircraft was at least 1000 feet too high to intercept the glide slope. In both cases the text accurately listed a series of actions that created the infeasible procedure.

The two structure variants and two presence of rationale variants resulted in four distinct display formats. The structure was either sequential or concurrent. The sequential structure listed all the actions that were required to complete the descent in a single column and noted when to do each action by attaching a 'fix' to the action that was also presented on the graphical display (see Figure 1). The concurrent structure listed the actions in a matrix where the columns related to horizontal motion, vertical motion, speed, or configuration, with all concurrent actions listed in the same row (see Figure 2). Again each row was notated with a fix and/or event to indicate when they should be done. The rationales, when provided, explained why an action should be done in general and/or done at a particular time (see Figure 1). Thus, the four formats are sequential, sequential with rationales, concurrent, and concurrent with rationales.

We blocked on the factor rationale since it was possible that there would be some learning, so the pilots either saw a combination of scenarios 1-4 and then scenarios 5-8 or they saw scenarios 5-8 and then scenarios 1-4. We used 8 different scenario orders; four pilots did each order of scenarios (see Table 1).

Table 1: Scenario Descriptions

Scenario	Condition	Accuracy	Structure	<u>Rationale</u>
1	PA	Accurate	Sequential	Not present
2	PA	Inaccurate	Concurrent	Not present
3	NPA	Inaccurate	Sequential	Not present
4	NPA	Accurate	Concurrent	Not present
5	NPA	Inaccurate	Concurrent	Present
6	NPA	Accurate	Sequential	Present
7	PA	Accurate	Concurrent	Present
8	PA	Inaccurate	Sequential	Present

Measurements

Measurement consisted of the pilots' responses to the presented procedures and a follow-up questionnaire. The pilot procedure response measurements were the pilots' responses (i.e., would or would not follow the procedure), the confidence they assigned to their response, the correctness of their responses (i.e., whether their response matched the flight procedures' accuracy), and the correctness of the pilots' reasoning about the procedure as recorded in written comments. The questionnaire measurements are the pilots' opinions about the different presentations of the procedure.

Results

There were 256 data points for each of the response variables: pilot responses, pilot confidence, correctness of pilot responses, and correctness of pilot reasoning. In addition to the four experimental factors, the pilot group, which represents the order in which the pilots saw the different scenarios, was also examined for main effects, but was shown to not have an effect for any of the response variables.

Pilot Responses

An analysis of variance (ANOVA) general linear model (GLM) (type III adjusted sum of squares) was used as the analysis method. The GLM for pilot responses versus the four experimental factors: performance of the aircraft (PA or NPA), procedure accuracy (accurate or inaccurate), procedure structure (sequential or concurrent), and the presence of rationale showed that only the aircraft condition, PA or NPA, is a statistically significant factor ($p < 0.01$). Examination of the data shows that pilots were more likely to respond 'No' (not comfortable following) in performance-altered conditions and 'Yes' (comfortable) in non-performance-altered conditions.

Pilot Confidence Level in Response

The GLM for pilot confidence level versus the experimental design factors also had a significant factor – performance of the aircraft once again ($p < 0.05$). Examination of the data showed that the pilots had a higher level of confidence with non-performance altering conditions. This is not surprising but does show that the pilots did account for the aircraft performance when making their judgment.

Correctness of Pilot Response

For the correctness of the pilots' responses when compared to the accuracy of the presented procedures, none of the four experimental factors had a statistically significant effect. In fact, on the whole the pilots did little better than chance (52%) on correctly judging the accuracy of the presented procedures.

Correctness of Pilot Reasoning

Finally, the correctness of the pilots' reasoning for accepting or not accepting a procedure was analyzed by categorizing pilots' reasoning and then comparing these categorizations with those provided by a subject matter expert. Looking at the four experimental factors versus correctness in pilot reasoning showed that two of the factors were statistically significant: accuracy of the presented procedure ($p = 0.05$) and rationale ($p < 0.01$). Examination of the data showed that the pilots had more accurate reasoning for accurate scenarios. This is not surprising since they basically only had to agree that it was done correctly. In addition, further analysis showed that procedures displaying rationale resulted in more correct reasoning by the pilots.

Questionnaire Results

The questionnaire measures came from the opinions of the pilots on the four different formats. Of the four different formats, 45% of the pilots with an opinion preferred the

concurrent with rationale format. Overall, 67% of the pilots with an expressed opinion preferred the concurrent format over the sequential format, and 91% of the pilots with an expressed opinion preferred having rationales over not having rationales.

Discussion

There were large individual differences between the pilots in their acceptance of the flight procedures and their reasoning for that acceptance. In addition, overall the pilots were not any better than chance at distinguishing feasible procedures from infeasible procedures. This is not surprising since this is a task that they have rarely, if ever, performed and does not have any standardized training. Pilots do practice emergency situations in simulator training but these often focus on the initial procedural response to the emergency as opposed to generating a new flight procedure on the fly to descend to an airport.

However, there are several interesting aspects of the results of this study. Not only were the pilots more likely to accept a procedure in the NPA condition, they also had more confidence in that acceptance. This may be due to a higher level of comfort with a “normal” aircraft that should perform as expected. However, this comfort may be misapplied, as they often indicated they would follow inaccurate NPA flight procedures, and had greater confidence in their judgments.

When the four experimental factors were examined in relation to correctness of pilot reasoning, procedure accuracy and the presence of rationale were significant. Having the correct reasoning for an accurate procedure was not overly difficult since basically the pilot had to just accept the procedure as correct without listing caveats. More interestingly, the procedures with rationales lead to a more correct reasoning by the pilot for acceptance or non-acceptance of a procedure. The pilots also reported that they liked being provided with the rationale of a procedure.

There is no support for the structure impacting the pilots’ responses or correctness in the objective results, but a majority of the pilots did prefer the concurrent structure over the sequential structure.

Summary of Work to Date

This study suggests that the presence of rationales or explanations for automatically generated decisions can aid the operators in more correct reasoning about that decision; however, it did not impact the correctness of their response to follow or not follow the decision. Further investigation is needed to see why this contextual information did not also make the pilots’ judgment more correct. However, these results indicate that including rationale with a suggested plan of action can improve some aspects of operator performance which might lead to enhanced system function and help operators deal with the potential brittleness of automatic systems. This finding supports both the design of procedures and the design of automatic systems that suggest courses of action to a human in situations that cannot be fully evaluated by the procedure or automatic system.

On-Going Work

In the coming year, we plan to build on these results through two activities: the development of heuristics suitable for planning emergency descents; and a possible development of training material for pilots that incorporates these heuristics.

References

Billings, C. E. (1997). *Aviation Automation: The Search for a Human-Centered Approach*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

- Ockerman, J. J., & Pritchett, A. R. (2000). A Review and reappraisal of task guidance: Aiding workers in procedure following. *International Journal of Cognitive Ergonomics*, 4(3), 191-212.
- Pritchett, A. R., Nix, D. C., & Ockerman, J. J. (2001). *Empirical evaluations of pilot planning behavior in emergency situations*. Paper presented at the 4xth Annual Meeting of the Human Factors and Ergonomics Society, Minneapolis, MN.

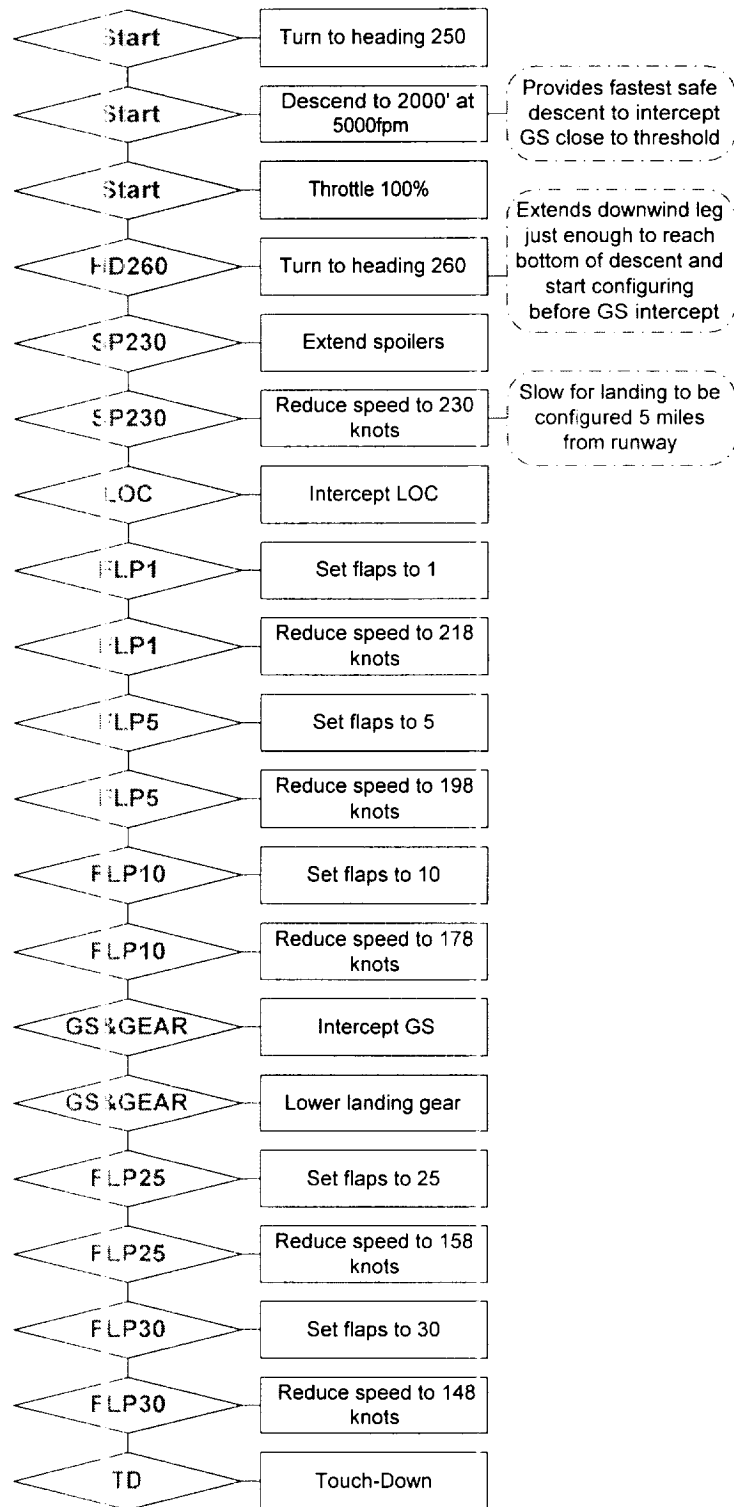


Figure 1: Sequential Structure with Rationale

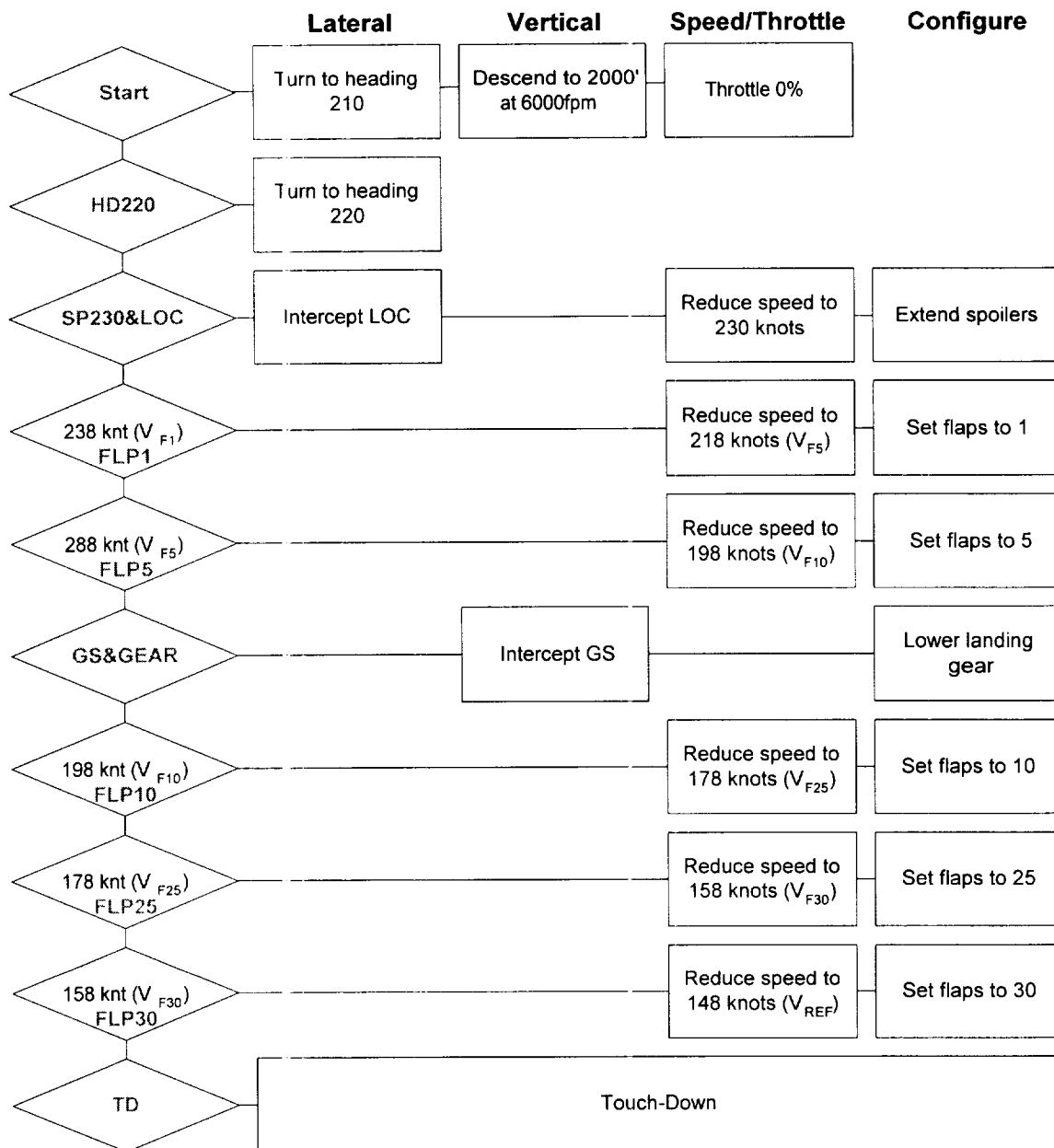


Figure 2: Concurrent Structure without Rationale